Great Learning
POWER AHEAD

Supervised Learning Regression Interview Questions

1. What is Linear regression? What is the difference between Simple and Multiple Linear Regression?
2. Explain mathematically how Linear Regression works.
3. What is meant by the Line of Best Fit?
4. What are the assumptions for applying linear regression?
5. What is Regularization? Explain the Bias Variance Tradeoff?
6. Why are Lasso Regression and Ridge Regression used?
7. How to perform feature selection, compensating for overfitting, and smoothing?
8. What is multicollinearity and why we should be very careful to take care of it before applying multiple regression?
9. What are the ways of checking for multicollinearity?
10. What are the ways of handling multicollinearity?
11. What metrics are used to evaluate a regression model?
12. What is R2 ? how does R2 compare with adjusted R2?
13. What is the coefficient of determination?
14. What is VIF and what is it used for in Linear Regression?
15. What is logistic regression and how does it work?
16. What is the difference between Correlation and Covariance?
17. Researchers believe that there is a correlation between smoking status and infertility. A sample of 30 females was tested for this purpose. Write the hypothesis for this scenario?
18. How do we check whether correlation coefficient, r is significant or not?
19. What is meant by explained variation? How is it computed?
20. What is meant by unexplained variation and how is it computed

21. How does a Non-Linear regression analysis differ from a Linear regression analysis?

22. How does the autocorrelation of Errors influence the Standard Error of the model? And what can be done to rectify that?

23. Is the vertical offset, horizontal offset, or the perpendicular offset minimized for least-square fitting, assuming that the vertical axis is the dependent variable? Why is this so?

24. What scenario would you prefer to use Gradient Descent instead of Ordinary Least Square Regression and why?

25. When the Line of Best Fit evolves, the distance between Actual and predicted values is a perpendicular distance - True or False and why?

26. If you observe that the test error is increasing after a certain number of iterations, what do you infer is most likely to be occurring? How do you address this problem?

27. What is the difference between Squaring of Lambda/alpha vs keeping the lambda absolute? How does it impact the model?

28. If the data is suffering from Multicollinearity, how do you think you can deal with that problem before building the model.

29. What are the differences between Linear Regression and Logistic Regression?

30. Can Logistic Regression be used for more than 2 classes?

31. How overfitting is controlled by introducing the Bias. Is there any other way we can reduce the overfitting of the model?

32. How do you remove variable redundancy?

33. Why normality of residuals is important in linear regression?

34. Why do we square the residuals instead of using modulus?

35. List down the techniques that are adopted to find the parameters of the linear regression line which best fits the model.

36. When should the Gradient Descent method be preferred instead of the Normal Equation in Linear Regression Algorithm?

37. Is it possible to apply Linear Regression for Time Series Analysis?

38. You run your regression on different subsets of your data, and in each subset, the beta value for a certain variable varies wildly. What could be the issue here?

39. Your linear regression doesn't run and communicates that there is an infinite number of best estimates for the regression coefficients. What could be wrong?

40. What is robust regression?